Infinite Data Generator & Experiment Analysis Tool for Human Pose Machines

Sabih Bin Wasi, Yaser Sheikh, Lei Tan

Motivation

Like most machine learning algorithms, the progress in human pose estimation systems have been hindered by **limited training data** and **ineffecient methods to identify the right feature set.**
The current largest dataset available for human pose machine is merely 10K annotated images - insufficient to learn multidimension vector for joints prediction. These datasets are small because **manual annotation is time-consuming, expensive and error-prone**.
On the other hand, the complexity of the problem makes **perfomance analysis of algorithm difficult and unstructured**.

Methodology

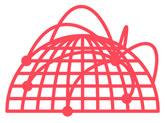
Our approach is to build automated data generation machine which can produce diverse set of annotated images. To achieve this, we capitalized on Panoptic Studio infrastructure at Robotics Institute.

Kinect Sensor
3D annotations for 25 body joints

+

Panoptic Studio
480 HD and VGA cameras
All Synchronized

=



Geometric Projections
12,000 images per second
Quick, Cheap and Accurate

Experiment

We captured 4 image sequences with different subjects inside the dome. The subjects were asked to perform 31 poses spanning common body movements. **The sequences were 90 seconds long which gave us 4.3M images.** We then sampled every 5th frame to capture the essence of sequence without overfeeding the dataset. The initialization, capture, post processing took 48 hours

48 hours **850K annotated images**

Design

We segmented our experiment analysis task into 3 sets:

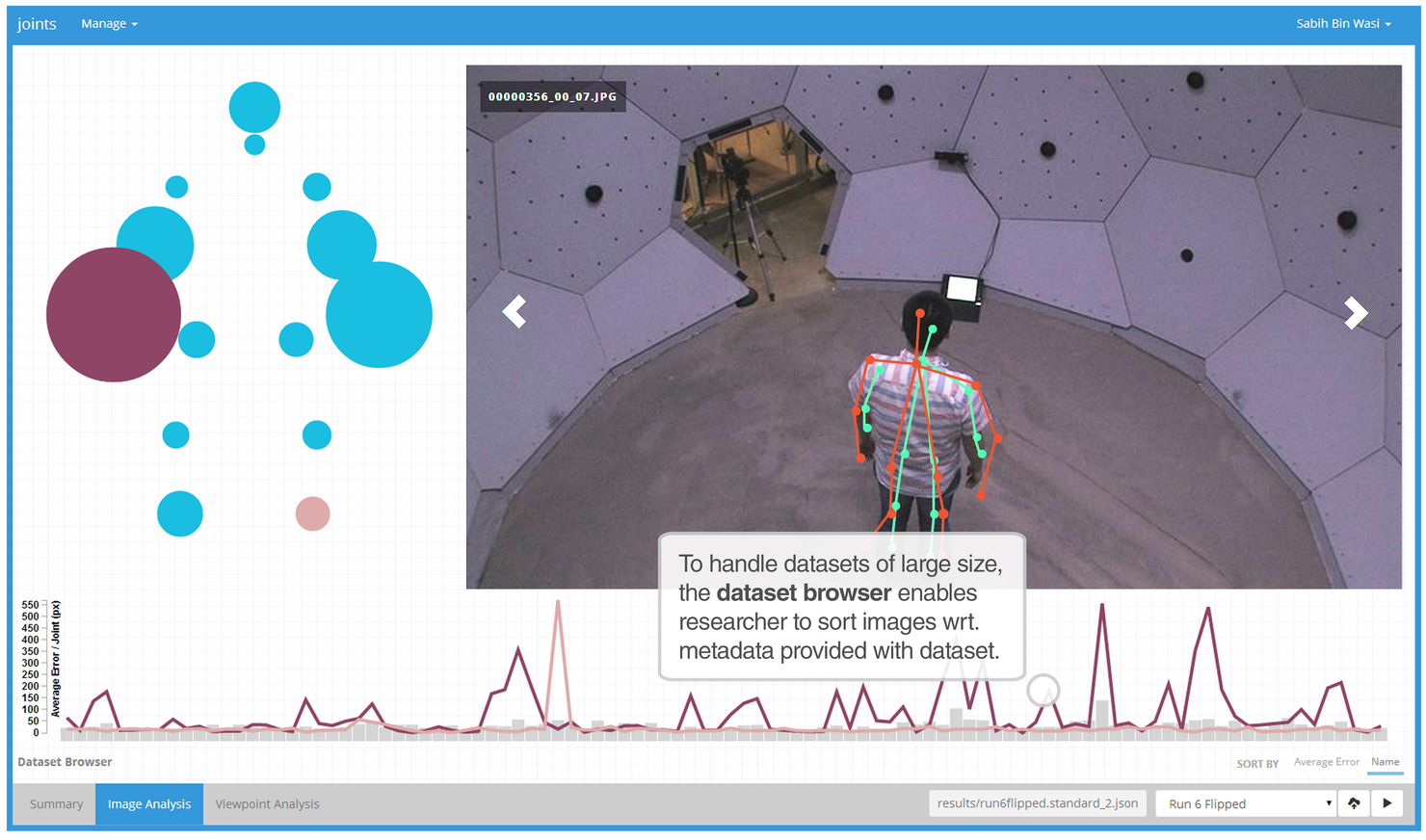
1. Aggregate-data Analysis: Through this section, researcher can quickly compare his system's performance with either other systems or different configuration of his own system.

2. Image-level Analysis: The tool draws skeletons recieved from multiple runs on an image.

3. Extensible Metadata Analysis: With metadata attatched to dataset (pose-id, viewpoint etc.), the tool can analyse system's performance over different values of such metadata field. This enables researchers to not just identify the featureset that performs differently under different image set but also help them build useful dataset.

Implementation

To ensure portability and **wide audience reach**, we use light AJAX web application framework.
To store data, we use JSON and MySQL server for **quick dataset retrieval.** The format of the datasets and test sets were designed to accomodate popular pose machine systems.
Fluid data visualization is programmed in d3js.



Conclusion

Using data-driven modular visualization and experiment analysis, our tool could help researchers developing human pose estimation systems in **building better training datasets**. Additionally, by **extensive, easier and visual comparison** of different configurations of their system, they can identify the right feature set and parameters for their machine.

Future Work

With aid from theatre language, we intend to identify the set of **human poses that encodes all possible body movements**. Also, a smarter way to pick the best representative images from large, repetitive dataset needs to be devised.
Ability to **add useful metadata with each captured image** would be useful for system analysis.

Further features on all three levels of analysis are required to aid researchers. T**hreshold-based aggregate analysis** can help them understand the usefulness of their system.
Image-level analysis can become more useful if **scoremaps for each joint** classification is displayed though that requires extensive test results.

Acknowledgements: Hanbyul Joo, Sanjib Banerjee, Varun Ramakrishna

Carnegie Mellon
THE ROBOTICS INSTITUTE